

## Aspekte digitaler Infrastrukturen

### Herausforderungen und Perspektiven von Forschungsdaten im Sonderforschungsbereich 980 „Episteme in Bewegung“

**Zusammenfassung:** Seit dem Jahr 2016 ist am Sonderforschungsbereich 980 „Episteme in Bewegung“ ein Informationsinfrastrukturprojekt angesiedelt, das fachwissenschaftliche und informationstechnische Expertise bündelt und eine zentrale Infrastruktur für alle Teilprojekte etabliert. Anhand konkreter Arbeitsschritte beim Umgang mit Forschungsdaten wird in diesem Beitrag gezeigt, welche Sichtweisen verschiedene Akteure – Sonderforschungsbereich, Fachwissenschaftler, IT, Informationseinrichtungen – einnehmen und wie diesen infrastrukturell begegnet wird.

**Schlüsselwörter:** Forschungsinfrastrukturen, Sonderforschungsbereiche

### **Aspects of Digital Infrastructures: Challenges and Perspectives of Research Data in the Collaborative Research Centre “Episteme in Motion”**

**Abstract:** Since 2016 the Collaborative Research Centre 980 “Episteme in Motion” includes an information infrastructure project which combines humanities and informatics expertise and establishes a central infrastructure for all participating projects. This article shows perspectives of affected stakeholders – the CRC, the humanities researcher, the IT, the information facilities – for different aspects of research data handling and how to encounter them from an infrastructure perspective.

**Keywords:** Research Infrastructures, Collaborative Research Centres

## 1 Einleitung

Der 2012 an der Freien Universität Berlin etablierte Sonderforschungsbereich 980 „Episteme in Bewegung. Wissenstransfer von der Alten Welt bis in die frühe Neuzeit“<sup>1</sup> untersucht Prozesse des Wissenswandels in europäischen und nichteuropäischen Kulturen der Vormoderne. In derzeit 25 Teilprojekten und über 80 Forschenden aus 27 Fachdisziplinen bündelt der von der Deutschen

---

<sup>1</sup> <http://www.sfb-episteme.de/>

Forschungsgemeinschaft (DFG) geförderte SFB vor allem die ‚Kleinen Fächer‘ der Freien Universität aus den Fachbereichen Philosophie und Geisteswissenschaften sowie Geschichts- und Kulturwissenschaften. Aus dieser disziplinären Vielfalt speist sich das breite Spektrum von Gegenständen – von der altägyptischen Pyramideninschrift bis hin zum frühneuzeitlichen Sprachlernbuch für Geschäftsreisende – und die interdisziplinäre Expertise, derer es bedarf, um historische Wissensbestände und -bewegungen aus transkultureller Perspektive zu untersuchen.

Seit Beginn der zweiten Förderphase 2016 unterhält der SFB zudem ein Informationsinfrastrukturprojekt (INF),<sup>2</sup> das eine digitale Dateninfrastruktur für alle Teilprojekte des SFBs entwickelt. Die entwickelten Methoden und Verfahren sollen möglichst projektübergreifend nutzbar sein und auch solche Forschungsfragen unterstützen, die künftige Vorhaben an das Material stellen. Das INF-Teilprojekt bündelt sowohl fachwissenschaftliche als auch informationstechnologische und infrastrukturelle Expertise, um einerseits innovative Forschungsmethoden der Digital Humanities zu entwickeln und zu erproben, andererseits Ergebnisse nachhaltig und nachnutzbar verfügbar zu machen.

Im Folgenden soll das Zusammenspiel der verschiedenen Akteure in einer digitalen Infrastruktur vor dem Hintergrund des geisteswissenschaftlichen Forschungsprozesses, der Generierung, Anreicherung und nachhaltigen Bereitstellung von Forschungsdaten in den Blick genommen werden. Im Rahmen des auf keinen historischen Gegenstand festgelegten INF-Projekts kann diese Fragestellung zur Selbstbeobachtung anregen und diese Selbstbeobachtung kann nach Akteuren oder Rollen von Akteuren differenziert werden:

1. Der Sonderforschungsbereich: Die Vertreter des Sonderforschungsbereichs sind über ein gemeinsames Erkenntnisinteresse miteinander verbunden, das in der Regel nicht primär auf eine digitale Infrastruktur gerichtet ist. Digitale Infrastrukturen bilden aber insofern einen zentralen Teil des Sonderforschungsbereichs, als sie ein methodisches Erweiterungspotential darstellen, um komplexe und verflochtene Arten von Wissenstransfers und der Wissensproduktion genauer erforschen zu können. Insofern werden digitale Infrastrukturen, die im INF-Projekt geschaffen werden, in die alltägliche Arbeits- und Forschungstätigkeit der Geisteswissenschaftler integriert.
2. Die Fachwissenschaftler: Fachwissenschaftler können diverse Erwartungen und Anforderungen unterschiedlichen Detailgrades an ein INF-Projekt haben. Es liegt auf der Hand, dass die Erwartungen auch von der Präsentation des INF-Projekts selbst beeinflusst werden können und der Erfolg der Integration des Projekts in den Sonderforschungsbereich auch von Präzisierungen der Erwartungen, Forschungsfragen und Anforderungen abhängt.
3. Die IT-Perspektive: Auf technischer Seite stehen im INF-Kontext der Aufbau einer Forschungsdateninfrastruktur und die dafür notwendige Softwareentwicklung im Fokus. Da an dieser Stelle neuartige Lösungen für die spezifischen fachwissenschaftlichen Problemstellungen

---

<sup>2</sup> <http://www.sfb-episteme.de/teilprojekte/informationsinfrastruktur/index.html>.

konzipiert werden, stehen neben Serviceleistungen und Betrieb auch Forschungsaspekte aus den Bereichen Informationswissenschaft und Digital Humanities im Mittelpunkt.

4. Die Informationseinrichtungen: Gemäß den Anforderungen der DFG, die eine Zusammenarbeit mit den einschlägig ausgewiesenen Informationseinrichtungen am Standort sowie die Nutzung bereits vorhandener Repositorien und Werkzeuge vorsehen,<sup>3</sup> erfolgt eine kontinuierliche Begleitung der Forschungsprozesse im Hinblick auf eine nachhaltige Sicherung der Projektergebnisse an der Freien Universität Berlin. Im Kontext des Forschungsdatenmanagements und der Forschungsdateninfrastruktur sollen die Ergebnisse des INF-Teilprojekts den Auf- und Ausbau nachhaltiger, institutionsübergreifender Strukturen an der Freien Universität unterstützen und für andere Forschungskontexte verfügbar gemacht werden. Es ist, wie gezeigt werden soll, von zentraler Bedeutung, dass Vertreter dieser einzelnen Perspektiven nicht nur auf die Ausrichtung eines Informationsinfrastrukturprojektes von außen einwirken, sondern vielmehr auch innerhalb des INF-Projektes zusammenkommen. Das heißt, dass die unterschiedlichen Perspektiven dieser Akteure und ihrer Rollen dort repräsentiert werden. Nur so ist es möglich, die Inhalte langfristig in Ausrichtung auf diese vielseitigen Interessen hin gemeinsam auszugestalten. Im Gegenzug bedeutet dies, dass die Zusammenarbeit im Projekt über interdisziplinäre Wissenschaft hinausgeht – fachwissenschaftliche und informationstechnische Forschung und Entwicklung treten in einen engen kollaborativen Zusammenhang, in dessen Umsetzung die fachwissenschaftliche und die informationstechnische Seite gleichermaßen und gleichwertig involviert sind. Im Bezug auf das dabei geleistete Forschungsdatenmanagement in INF-Projekten ist der Begriff des „embedded data manager“<sup>4</sup> diskutiert worden, als Rolle von Forschungsdatenexperten, die innerhalb des Projektkontextes, sozusagen lokal, den Forschungsprozess hinsichtlich des Lebenszykluses von Forschungsdaten unterstützt. „In dem noch nicht abgesteckten Arbeitsfeld werden Data Manager allerdings häufiger auch auf technische Tätigkeiten reduziert und als ‚Datenklempner‘ charakterisiert.“<sup>5</sup> Diese Zuweisung ist insbesondere dann von der Hand zu weisen, wenn im INF-Projekt technische Partner aus dem wissenschaftlichen Bereich neben die fachwissenschaftliche Seite des Sonderforschungsbereichs treten. Vielmehr kann die eng verzahnte Forschung und Entwicklung dann zum Nukleus für innovative Ergebnisse auf inhaltlicher wie auf technischer Seite werden.

Insgesamt treten im Rahmen des INF-Projekts drei Prozesslogiken nebeneinander. Erstens stehen die ‚klassischen‘ Forschungsprozesse im Mittelpunkt. Im Rahmen des Sonderforschungsbereichs 980 sind dabei traditionell arbeitende geisteswissenschaftliche Disziplinen versammelt, die sich nicht unbedingt als datengetrieben verstehen. Zum zweiten zeichnen sich INF-Projekte verantwortlich für die Abwicklung

---

<sup>3</sup> Deutsche Forschungsgemeinschaft (2018) 11.

<sup>4</sup> Cremer, Engelhardt und Neuroth (2015).

<sup>5</sup> Cremer, Engelhardt und Neuroth (2015) 17.

von Prozessen im Sinne des sog. *research data lifecycle*: Erzeugung, Speicherung, Nutzung und Langzeitsicherung von Forschungsdaten. Drittens treten Prozesse der Softwareentwicklung hinzu, insofern Infrastrukturkomponenten und Werkzeuge adaptiert oder neu entwickelt werden. Dies geschieht aufgrund der beschriebenen Projektstruktur kaum in Form klassischer Prozessmodelle (Wasserfallmodell, etc.), sondern in iterativer Zusammenarbeit zwischen Entwickler und Nutzer. Die Verzahnung der verschiedenen Prozesse und Perspektiven erzwingt auf allen Seiten einen organisatorischen und kommunikativen Mehraufwand. In diesem Sinne sind Infrastrukturprojekte angesichts der Heterogenität, Spezifität und Menge von Forschungsinteressen in einem Forschungsvorhaben in dieser Form nur leistbar, wenn alle beteiligten Seiten langfristig Ressourcen dafür zur Verfügung stellen. Ideal eignen sich ‚Pilotprojekte‘, in denen neue Konzepte und Werkzeuge in vertiefter, modellhafter Zusammenarbeit ausgearbeitet und erprobt werden. In solchen sogenannten Pilotprojekten werden ausgehend von einer konkreten Forschungsfrage eine digitale Infrastruktur und an diese angeschlossene Tools entwickelt, die für die Arbeit an der jeweils konkreten Forschungsfrage des Projekts eine methodische Erweiterung bedeuten und einen Mehrwert darstellen. Ohne die neue Infrastruktur wäre die Forschungsfrage in der Regel nicht, nur sehr schwer oder deutlich umständlicher zu lösen. Im Falle des Sonderforschungsbereichs 980 sind konkret Projekte als Pilotprojekte ausgewählt worden, an denen sich besonders gut digitale Tools entwickeln ließen, mit deren Hilfe so genannte innere Transferprozesse (d.h. Transfers innerhalb von Wissensbeständen – etwa innerhalb von Aristoteleshandschriften oder frühneuzeitlichen Sprachlehrbüchern) und äußere Transferprozesse (d.h. Transfers in Raum und Zeit – z.B. der raumzeitlichen Transfers bestimmter Sargtexte im Alten Ägypten) sowie deren jeweilige Multidirektionalität besser erforscht werden können. Ziel war und ist, die Projekte bei der Beantwortung ihrer Forschungsfragen möglichst direkt zu unterstützen und gleichzeitig aufgrund ihrer disziplinären Heterogenität verallgemeinerbare Lösungen zu erarbeiten.

## **2 Informationsinfrastruktur für die Erforschung von Wissenstransfers**

Der Sonderforschungsbereich untersucht Wissenstransfers in der Zeit von der ‚Alten Welt‘ bis zur Frühen Neuzeit. Mit Blick auf die Pilotprojekte ließen sich Kategorien solcher Prozesse bestimmen, die für mehr als eines dieser Projekte relevant sind: räumliche und zeitliche Bewegung von Objekten und Personen, Wanderungen von Wissen und Veränderungen in gedruckten Büchern und Handschriften (kodikologische und typographische Transfers) sowie die damit jeweils verbundenen epistemischen Transfers. Diese gemeinsamen Kategorien erlauben idealerweise auch Übertragungen von Datenmodellen, digitaler Analyse und Visualisierung im Rahmen des Informationsinfrastrukturprojektes.

Am konkreten Beispiel eines einzelnen solchen Teilprojektes zu Aristoteles' *de interpretatione* lässt sich eine Ausgestaltung dieser Transfers veranschaulichen. Von besonderem Interesse ist in dem Vorhaben zu den etwa 150 griechischsprachigen Handschriften von *de interpretatione* das Verhältnis von kommentiertem Text, Scholien, Glossen und Diagrammen auf einer Handschriftenseite als einem Mittel der ‚Erkenntnissteuerung und Wissensverwaltung‘,<sup>6</sup> aber auch als stemmatologisches Indiz. Ergänzend zu Handschriftenbeschreibungen und zu diesem aristotelischen Text wurden hier auch ‚Anmerkungen‘ textkritisch berücksichtigt, die Erkenntnisse über diese als komplex geltenden Überlieferung eines in verschiedenen historischen Kontexten zentralen Lehrtextes versprechen. Gefragt wird, von welchem Autoren, aus welchem Kontext die Anmerkungen stammen und welche Ideen, welche Argumentationen in ihnen transportiert werden, oder auch ob das Argument von einem Diagramm richtig oder falsch abgebildet wird. Übereinstimmungen in hinzugefügten Diagrammen und Glossen können erlauben, die Position von Handschriften in den Zweigen der Überlieferung zu prüfen und vielleicht sicherer zu bestimmen. Angenommen werden kann, dass auch räumliche Transfers, insofern sie als ‚Neukontextualisierungen‘ von Wissens-elementen gedacht werden, innere Transfers in Wissensbeständen begründen. Über solche Neukontextualisierungen können schließlich Verwandtschaften von Handschriftengruppen erforscht werden, die wiederum in einzelnen Fällen Rückschlüsse auf einen ganz konkreten räumlichen und zeitlichen Transfer, auf Traditionen und Schulbildungen ermöglichen.

Das Konzept der Transfers bestimmt die Auswahl der zur Verfügung gestellten Bearbeitungs- und Analysewerkzeuge. Die Wissenstransfers bedingen unter anderem eine methodische Ausrichtung auf räumlich-zeitliche Analysen und Netzwerkanalysen. Dabei wird standardisierten Formaten (z.B. Text Encoding Initiative (TEI),<sup>7</sup> CIDOC CRM,<sup>8</sup> Web Annotation Data Model<sup>9</sup>) und etablierten Werkzeugen (z.B. Place and Time Navigator (PLATIN),<sup>10</sup> Software Workflow for the automatic tagging of medieval manuscript images (SWATI)<sup>11</sup>) der Vorzug gegeben gegenüber Neuentwicklungen, wo immer dies möglich ist. Dies stellt einen unverzichtbaren Baustein dar für Interoperabilität und *community building*

---

<sup>6</sup> Diese Begrifflichkeit findet sich in Anwendung auf die „Gestaltungsmittel [...], welche der Buchdruck zwischen ca. 1470 und 1660 nach und nach entwickelte“ in Enenkel (2005) 52–53. Vgl. auch Neuber (2000) 191: „Eine Seite ist immer zugleich ein ‚Bild‘, das heißt, das Layout hat seine spezifische kognitive Funktion, die die Lektüre, das Verinnerlichen und das Verstehen des Textes steuert.“

<sup>7</sup> <https://tei-c.org/>.

<sup>8</sup> <http://www.cidoc-crm.org/>.

<sup>9</sup> Young, Ciccarese und Sanderson (2017).

<sup>10</sup> <http://platin.mpiwg-berlin.mpg.de/>.

<sup>11</sup> Chandna et al. (2015).

über die Grenzen des Forschungsverbunds hinaus. Überdies ist es dafür unabdingbar, existierende digitale Wissensbestände nachzunutzen, beispielsweise Gazetteers, Normdaten (räumliche und zeitliche Transfers) und domainspezifische Vokabulare.

### 3 Forschungsdaten und Datenmodellierung

Aus der Perspektive der Deutschen Forschungsgemeinschaft werden Forschungsdaten im Kontext von SFB-Infrastrukturprojekten einerseits als Objekte eines Forschungsprozesses und andererseits als Resultate davon definiert.<sup>12</sup> Im Folgenden soll es dabei zunächst um (Meta)daten in ersterem Sinne gehen. Metadaten, die aus den Forschungsprozessen selbst entstehen, folgen in Abschnitt 5. Daten als Forschungsobjekte sind dabei erst einmal so heterogen wie die Forschungsprojekte selbst. Dies gilt insbesondere, da der Sonderforschungsbereich manchmal nicht selbst die datenerhebende Institution ist, sondern hier Daten unter einem spezifischen Forschungsinteresse aggregiert werden. So liegen die Daten in verschiedensten Formen und Formaten vor – nicht immer nach etablierten Standards oder mit Fokus auf Maschinenlesbarkeit. Im Kern des Interesses stehen in den Pilotprojekten Bild- und Textdaten, wobei Bilddaten von Digitalisaten häufiger vertreten sind als transkribierte Texte. Hinzu treten bibliographische, paläographische, kodikologische oder anderweitige Metadaten, die insbesondere gesichertes Wissen im Sinne eines bestehenden Forschungsstandes abbilden. Im SFB 980 verstärkt sich die Dichotomie von Informationen, die einen bestehenden status quo abbilden, und solchen, die darauf aufbauend originäre Forschungsergebnisse transportieren, zusätzlich durch einen Beginn des INF-Projektes erst in der zweiten Förderphase, also mit vier Jahren Versatz zur sonstigen Forschungsarbeit im SFB. Es entsteht so naturgemäß eine gewisse Grenze zwischen Datenerhebung 'vor INF' und 'nach INF'. Dieser vergleichsweise späte Aufbau eines Forschungsdatenmanagements sowie der zugehörigen Infrastruktur spiegelt dabei eine auch anderweitig häufig ausgeprägte Praxis wieder:

„Die meisten Fachwissenschaftler und andere Beteiligte beginnen mit Überlegungen zum Forschungsdatenmanagement erst kurz vor oder zum Teil auch erst nach der Förderphase ihres Projektes, wenn es gilt, die Mindestanforderungen zu erfüllen und für die Forschungsdaten einen geeigneten Speicherort, wie ein Datenzentrum oder Repository, zu suchen.“<sup>13</sup>

---

<sup>12</sup> „Research data [...] includes anything that is either the object of a research process or the result of it“. Redöhl (2016) 78f.

<sup>13</sup>Cremer, Engelhardt und Neuroth (2015) 15.

Die verbreitete Einstufung von Datenmanagement als abschließender Forschungsaktivität<sup>14</sup> läuft dem Konzept gemeinsamer Forschung und Entwicklung zuwider. Auch wenn der Forschungsaspekt bei ‚reinem‘ Forschungsdatenmanagement nicht auf den ersten Blick evident sein mag, so gewinnt er an Bedeutung, sobald man Forschungsdatenmanagement als „Zwilling von Konzepten wie E-Science, E-Research, Digital Humanities“<sup>15</sup> begreift – überall dort, wo Forschungsdaten nachhaltig gespeichert und verfügbar sein sollen sowie neuartige Konzepte, digitale Werkzeuge und Verfahren zur adäquaten Unterstützung von Forschungsfragen erst entstehen.

Sollen digitale Verfahren in die jeweilige Forschungsprozesse eingegliedert werden, ist eine nicht nur menschen- sondern auch maschinenlesbare Modellierung notwendig. Die Erhebung von kodikologischen Metadaten weist bereits auf die fachwissenschaftliche Komponente von digitalen Infrastrukturen, die neben die technische tritt.<sup>16</sup> Im Fall der Handschriftenbeschreibungen verfolgen sie den Zweck, Daten, die vorher nicht vollständig digital vorlagen, in einem standardisierten Format wie XML gemäß TEI P5 zu archivieren und Verweise auf Normdaten zu Orten und Personen zu ergänzen. Die ‚Transformation‘ ist aber nicht nur eine technische Frage. Das gewählte Format ist als Schema zugleich ein Modell des erfassten Gegenstands. Bei TEI P5 ist dies besonders deutlich, weil oft verschiedene Wege möglich sind, Angaben abzulegen. Dieser Prozess ist demnach auch eine fachwissenschaftliche Entscheidung. Die Datenmodellierung stellt damit eine Schicht dar, die technische und fachwissenschaftliche Komponenten der Infrastruktur verknüpft und sie kann Auswirkung auf die fachwissenschaftliche Sicht selbst haben. Die Art, wie Wissen gespeichert wird, bestimmt zu einem gewissen Grad, welche anderen unter anderem digitale Verfahren der Analyse, der Referenzierung, der Visualisierung und der Kontextualisierung sich anschließen können und welche nicht. Historische Daten sind dabei selten exakt und lückenlos. So kann es sich beispielsweise bei der Datierung von Handschriften in vielen Fällen nur um gröbere Einordnungen handeln, die aufgrund paläographischer oder kodikologischer Eigenheiten vorgenommen werden, wenn z. B. konkrete Datierungen in den Handschriften fehlen. Möglicherweise sind diese auch nur zu einer Seite hin (post, ante) begrenzt. Es können Indizien auf unterschiedliche Ergebnisse deuten und manche Bestimmung muss ausbleiben.

Diese Unschärfen und Lücken und auch der sich dabei ergebende Anteil an Interpretation und Schlussfolgerungen stellen eine Herausforderungen nicht nur für die Modellierung, sondern

---

<sup>14</sup> „Despite a growing awareness for data management, there is still a predominantly traditional perception of data management as a concluding activity of research endeavors in many scientific domains“, so Wang, Göpfert und Stark (2016) 15.

<sup>15</sup> Klump und Ludwig (2012) 262.

<sup>16</sup>Diese Bezeichnung ist von Neuroth und Rapp (2016) 267 übernommen, die darunter „Lehre und Forschung“ fassen. Im Folgenden wird der Aspekt der Lehre keine Erwähnung finden. Seine Bedeutung soll damit jedoch nicht geschmälert werden.

insbesondere auch für die Auswertung dar (siehe auch Abschnitt 6). Eine umfassende Maschinenlesbarkeit ist an diesen Stellen utopisch, insbesondere wenn auf detailreiche Darstellung gerade von Unschärfen nicht verzichtet werden kann und soll. Umso wichtiger ist es deshalb, technische und fachwissenschaftliche Seite von vornherein in einen engen Dialog zu bringen, da nur so Lösungen gefunden werden können, die bestmögliche Auswertbarkeit mit fachwissenschaftlicher Tiefe vereinen.

#### **4 Forschungsdatenrepositorium**

Das Kernelement der technischen Infrastruktur des Sonderforschungsbereiches ist nicht unerwarteterweise ein Forschungsdatenrepositorium, in dem Daten inklusive administrativer und beschreibender Metadaten aller Teilprojekte strukturiert verwaltet, nachhaltig gesichert und für weitere Forschung zugänglich gemacht werden. Im speziellen Fall dieses SFBs wird ein nicht unerheblicher Teil der Forschungsdaten wie beispielsweise Manuskriptdigitalisate oder Ausgrabungsdokumente nicht selbst erhoben bzw. generiert. Vielmehr agiert der SFB, wie bereits angesprochen, als ‚Datenaggregator‘, der die oftmals sehr verstreuten und schwer zugänglichen Einzeldatensätze sammelt und darauf aufbauend erschließt und anreichert, um Wissensbewegungen und -transfers aufzuzeigen. Jeder einzelne dieser Schritte kann dabei eine nicht unerhebliche Forschungsleistung darstellen, deren Ergebnisse nachvollziehbarerweise nachhaltig verfügbar und somit für die Fachcommunity nachnutzbar sein sollten.

Die Anforderungen der einzelnen Forscher an ein Datenrepositorium zeichnen sich, bereits vor dem Hintergrund der Vielfalt an im SFB vertretenen Fachdisziplinen, durch eine hohe Diversität aus. In der Zusammenarbeit kristallisieren sich jedoch in der Regel Anforderungen heraus, die eher eine Methodik oder Arbeitsweise betreffen als die eigentlichen Forschungsgegenstände. So sollen Daten oft ortsunabhängig erzeugt, durchsucht und bearbeitet werden können. Manchmal findet sich das Bedürfnis oder sogar die Notwendigkeit, diese Schritte kollaborativ, aber in einem geschützten Raum zu tätigen. Nicht selten müssen Daten aus verschiedenen Quellen aggregiert und stabil referenziert werden können. Diese letzte Forderung betrifft auch selbst erhobene oder generierte Daten, wenn sie publiziert werden sollen. Häufig sind aber auch Abstufungen in der Stabilität der Daten unumgänglich, weil sie während der Erstellung ‚dynamisch‘ bleiben müssen. Bedenkt man, dass auf Daten referiert werden kann, die sich selbst noch im Laufe des Projekts verändern können, wird die technische Anforderung deutlich, die dies bedeuten kann.



Die technische, organisatorische und strukturelle Nachhaltigkeit solcher Forschung und Entwicklung, wie sie im Kontext des INF-Projekts geleistet wird, ist Gegenstand strategischer Überlegungen und Konzepte, die über den Rahmen der unmittelbaren Förderung durch die DFG hinausgehen sollten. Im Zentrum stehen dabei Lösungsmodelle, die – derzeit selbst noch in der Entwicklung – auch nach Ende der Projektlaufzeit zur Verfügung stehen, die aber bereits heute in ihren generischen Angeboten so mitgestaltet werden können, dass sie den projektspezifischen Anforderungen gerecht werden. So ist nach Ende der Projektlaufzeit eine Speicherung und Bereitstellung der (rechtfreien) Daten in interoperablen Standardformaten im seit 2018 im Aufbau befindlichen institutionellen Repositorium der Freien Universität Berlin ‚ReFUBium‘<sup>17</sup> vorgesehen, das – u.a. durch den Nachweis in einschlägigen nationalen und internationalen bibliographischen Referenzsystemen sowie durch die Versehung mit persistenten Identifikatoren (DOIs) – nicht nur die nachhaltige Sichtbarkeit der im Forschungsverbund generierten Daten erhöht, sondern auch Folgeforschungen ermöglicht.

## **5 Forschungsprozesse und Datenanreicherung**

Obwohl es Ziel ist, die einzelnen Pilotprojekte möglichst ‚nah‘ zu begleiten, sollten auch Modelle, Verfahren und Werkzeuge möglichst nicht auf eine einzelne Forschungsfrage begrenzt bleiben. Annotationen als Verfahren, verschiedenste historische Quellen mit Wissen ‚anzureichern‘, haben sich in diesem Fall als Bindeglied über die Grenzen von Projekten und Disziplinen herausgestellt.

In vielen Fällen kann diese Anreicherung durch algorithmische Verfahren unterstützt werden, etwa durch eine Layoutanalyse von Manuskripten oder Drucken, die verschiedene Merkmale reproduzierbar vermisst. Aufgrund von beispielsweise schlecht erhaltenen Originalen, Digitalisaten geringer Auflösung oder komplexer Layoutstrukturen entsteht jedoch stets die Notwendigkeit, eine Korrektur und Evaluation der Ergebnisse zu ermöglichen. Um die gewonnenen Erkenntnisse der Forschung zugänglich zu machen und eine (technische) Nachnutzung zu erlauben, werden sie als digitale Annotationen in einer Annotationsinfrastruktur basierend auf den W3C-Empfehlungen Web Annotation Data Model<sup>18</sup> und Web Annotation Protocol<sup>19</sup> abgelegt. Ebenso werden auf diese Weise zusätzliche, strukturierte Informationen aus der Linked Open Data Cloud<sup>20</sup> zugänglich, sodass sich gleichermaßen spannende fachwissenschaftliche wie informatische Forschungsfragen ergeben.

---

<sup>17</sup> <https://refubium.fu-berlin.de>.

<sup>18</sup> Young, Ciccarese und Sanderson (2017).

<sup>19</sup> Sanderson (2017).

<sup>20</sup> <https://lod-cloud.net>.

Im konkreten Falle der Untersuchung von *de interpretatione* eröffnen digitale Annotationen etwa die Möglichkeit, Randerklärungen und erklärende Diagramme zu transkribieren, ihre Herkunft, soweit dies im Einzelfall möglich ist, zu kennzeichnen, sodass zum Beispiel Wanderungen einzelner Erklärungen und Diagramme von einer in die nächste Handschrift, von einem zu einem anderen Ort nachvollziehbar werden. Stellen können mit projektspezifischen Begriffslisten oder kontrollierten Vokabularen klassifiziert werden und müssen dabei auf unterschiedlichen Ebenen annotierbar sein. Eine Stelle kann zugleich ein Argument enthalten, eine Parallelstelle anführen und ein Indiz für eine historische Einordnung sein.<sup>21</sup>

Für die Untersuchung von Überlieferungs- und Wissensgeschichten ist dabei eine Klassifikation von annotiertem Objekt und Annotationszweck erforderlich. Auf diese Weise können sowohl beispielsweise ein Diagramm einer bestimmten, endlichen Menge von Diagrammtypen zugeordnet als auch Annotationen voneinander unterschieden werden, die Umschriften, Übersetzungen oder eben Informationen zum ‚Typ‘ liefern. Dabei ist nicht ausgeschlossen, dass sich diese Klassifikationen im Laufe des Forschungsprozess noch verändern. Trotz alledem stellen sie eine essentielle Voraussetzung für eine anschließende quantitative Auswertung dar, die dem Forscher nach dem (zeit-)aufwändigen Annotationsprozess Erkenntnisse verspricht.

## **6 Datenanalyse und Forschungsergebnisse**

Wie im gerade beschriebenen Beispiel der digitalen Annotationen können eine Datenanreicherung in den Forschungsprozess selbst integriert und die gewonnenen Informationen systematisch ausgewertet werden. Im Falle von *de interpretatione* kann in Kombination mit den Handschriftenbeschreibungen und ihrer Verknüpfung mit Normdaten zu Orten und Personen zum Beispiel eine Handschrift in einem überlieferungsgeschichtlichen Zweig positioniert werden, in dem man sie, ginge man nur vom aristotelischen Text aus, kaum verorten würde. Auf diese Weise können Annotationen an Bildern einerseits textkritisch genutzt und andererseits auch epistemische Transfers sichtbar werden.

Wie in jedem Forschungsprojekt hat die sichtbare Publikation solcher Ergebnisse inklusive der zugrundegelegten Daten im Sonderforschungsbereich eine besondere Bedeutung. Die angesprochene Aggregationen von Datenquellen und die Referenz von digitalen Annotationen können hier jedoch ein

---

<sup>21</sup> Vgl. die Überlegung in Jackson (2005) 137–138: „Marginalia record the responses of actual readers. If it were possible to assemble a reasonable number of annotated book from a particular period, we should be able to find out how each reader approached and reacted to the book at hand, and eventually to build up a collective impression.“ Im hier beschriebenen Projekt werden inhaltliche Überschneidungen gerade als Indizien für Gruppen untersucht.

Problem aufwerfen, wenn die referenzierten Datenquellen von Annotationen zum Beispiel aus lizenzrechtlichen Gründen bei Digitalisaten von Handschriftenseiten selbst nicht verfügbar sind oder nicht veröffentlicht werden können. Als problematisch lässt sich ebenso herausstellen, dass die betrachteten Forschungsdaten oftmals von Unschärfen und Lücken geprägt sind und damit belastbare quantitative Auswertungen deutlich erschweren. Es stellt sich daher im technischen Bereich die Frage, ob sich Methodiken und Algorithmen aus anderen Anwendungsfällen übertragen und ggf. anpassen lassen oder ob diese als Forschungsleistung gänzlich neu entwickelt werden müssen. Beispielhaft seien hier die Schwierigkeiten von verbreiteten Kollationswerkzeugen mit fragmentierten Texten oder die Verortung von Objekten unklarer Herkunft auf einer Karte zu nennen.

Die Frage, ob, von wem und wie weit solch angepasste oder entwickelte Forschungssoftware nachhaltig und langfristig betrieben werden kann und soll, ist zunehmend Gegenstand aktueller Debatten.<sup>22</sup> Auch für den SFB kann sie noch nicht hinreichend beantwortet werden: im Kontext der Reproduzierbarkeit von Forschungsergebnissen wird das Votum für eine langfristige Bereitstellung bei spezifischen Algorithmen einer Layoutanalyse eher positiv, bei einer graphische Annotationsoberfläche, die Informationen standardkonform ablegt, eher negativ ausfallen. Auch fließen hier aktuelle Diskussionen und Angebote großer, geisteswissenschaftlicher Verbundinfrastrukturen wie DARIAH-DE<sup>23</sup> in die Überlegungen mit ein. Die Bereitstellung und Publikation der im INF-Projekt entwickelten Software im Infrastrukturangebot von DARIAH-DE bietet die Einbettung der Forschungsergebnisse in das Serviceangebot einer etablierten Forschungsinfrastruktur, die die Anschlussfähigkeit auch an die aktuellen Entwicklungen auf nationaler und internationaler Ebene gewährleistet und die bundesweiten Diskussionen und Überlegungen zum nachhaltigen Forschungsdatenmanagement in den Geisteswissenschaften wesentlich mitgestaltet.

## **7 Schlussfolgerungen**

Dargestellt wurde die Konzeption eines Informationsinfrastrukturprojekts, das nicht nur ein Repositorium und allgemeine Werkzeuge für einen Sonderforschungsprojekt zur Verfügung stellen, sondern auch den Forschungsprozess einzelner Teilprojekte möglichst unmittelbar digital unterstützen möchte. Zentrale Herausforderungen sind die Heterogenität der konkreten Daten, Arbeitspraktiken und Forschungsfragen

---

<sup>22</sup> Im geisteswissenschaftlichen Bereich so z.B. die 2018 gegründete Arbeitsgemeinschaft „Research Software Engineering“ des Verbands Digital Humanities im deutschsprachigen Raum (DHd): <http://dig-hum.de/ag-research-software-engineering-digital-humanities>.

<sup>23</sup> <https://de.dariah.eu>.

sowie die nicht widersprüchlichen, aber je etwas eigenen Interessen des Sonderforschungsbereichs, der Forscher, der IT und der Informationseinrichtungen. Schließlich kann für die geisteswissenschaftliche Forschung in der Regel kein Modell unverändert angewendet werden wie bei Forschungen, die auf großen Mengen maschinell erhobener, ‚uninterpretierter‘ Daten beruhen. Da viele Schritte im Forschungsprozess, von der Datenmodellierung über die Adaption digitaler Werkzeuge bis zur Auswertung, sowohl eine informatische als auch eine fachwissenschaftliche Seite besitzen, ist eine besonders enge Zusammenarbeit vonnöten, bei der Geisteswissenschaftler in das INF-Projekt und Datenexperten in die Teilprojekte eingebunden werden. Wenn digitale Methoden möglichst unmittelbar in geisteswissenschaftliche Forschungsprozesse integriert werden sollen, dann ist eine institutionelle und personelle Einbettung, wie sie hier vorgestellt wurde, unumgänglich.

## Literaturverzeichnis

- Chandna, Swati; Tonne, Danah; Jejkal, Thomas; Stotzka, Rainer; Krause, Celia; Vanscheidt, Philipp; Busch, Hannah; Prabhune, Ajinkya (2015): Software workflow for the automatic tagging of medieval manuscript images (SWATI). In: *Document Recognition and Retrieval XXII. International Society for Optics and Photonics*, 940206.
- Cremer, Fabian; Engelhardt, Claudia; Neuroth, Heike (2015): Embedded Data Manager – Integriertes Forschungsdatenmanagement: Praxis, Perspektiven und Potentiale. In: *BIBLIOTHEK – Forschung und Praxis*, 39 (1), 13–31. DOI: 10.1515/bfp-2015-0006.
- Deutsche Forschungsgemeinschaft (DFG) (2018): Merkblatt Sonderforschungsbereiche. DFG-Vordruck 50.06. URL: 07/18 [http://www.dfg.de/formulare/50\\_06/50\\_06\\_de.pdf](http://www.dfg.de/formulare/50_06/50_06_de.pdf).
- Enenkel, Karl. A. E. (2005): *ARS ANTIQUITATIS*: Erkenntnissteuerung und Wissensverwaltung in Werken zur römischen Kulturgeschichte (ca. 1500–1750). In: Enenkel, Karl A. E.; Neuber, Wolfgang (Hrsg.): *Cognition and the Book. Typologies of Formal Organisation of Knowledge in the Printed Book of the Early Modern Period*. Leiden: Brill (Intersections 4), 51–123.
- Jackson, H. J. (2005): ‘Marginal Frivolities’: Readers’ Notes as Evidence for the History of Reading. In: Myers, Robin; Harris, Michael; Mandelbrote, Giles (Hrsg.): *Owners, Annotators and the Signs of Reading*. New Castle: Oak Knoll Press, 137–51.
- Klump, Jens; Ludwig, Jens (2012): Forschungsdaten-Management. In: Neuroth, Heike; Lossau, Norbert; Rapp, Andrea (Hrsg.): *Evolution der Informationsinfrastruktur*. Boizenburg: Verlag Werner Hülsbusch, 257–75.
- Neuber, Wolfgang (2000): Ökonomien des Verstehens. Markt, Buch und Erkenntnis im technischen Medienwandel der Frühen Neuzeit. In: Wenzel, Horst; Seipel, Wilfried; Wunberg, Gotthart (Hrsg.): *Die Verschriftlichung der Welt. Bild, Text und Zahl in der Kultur des Mittelalters und der Frühen Neuzeit*. Wien: Kunsthistorisches Museum Wien (Schriften des Kunsthistorischen Museums, 5), 181–211.
- Neuroth, Heike; Rapp, Andrea (2016): Nachhaltigkeit von digitalen Infrastrukturen. In: *BIBLIOTHEK – Forschung und Praxis*, 40 (2) 264–70.
- Redöhl, Brit (2016): The DFG Perspective: Research Data Management with a Focus on Collaborative Research Centres (SFB). In: Curdt, Constanze; Willmes, Christian (Hrsg.): *Proceedings of the 2nd Data*

*Management Workshop*. Köln: Geographisches Institut der Universität zu Köln – Kölner Geographische Arbeiten, 77–81. DOI: 10.5880/TR32DB.KGA96.12

Sanderson, Robert, Web Annotation Protocol. W3C Recommendation, 23. Februar 2017. Verfügbar unter <https://www.w3.org/TR/2017/REC-annotation-protocol-20170223>.

Wang, Wei Min; Göpfert, Tobias; Stark, Rainer (2016): Data Management in Collaborative Interdisciplinary Research Projects – Conclusions from the Digitalization of Research in Sustainable Manufacturing. In: *ISPRS Int. J. Geo-Inf.*, 5 (4) Art. 41.

Young, Benjamin; Ciccarese, Paolo; Sanderson, Robert: Web Annotation Data Model. W3C Recommendation, 23. Februar 2017. Verfügbar unter <https://www.w3.org/TR/2017/REC-annotation-model-20170223>.



Germaine Götzelmann  
Karlsruher Institut für Technologie  
Steinbuch Centre for Computing, Hermann-von-Helmholtz-Platz 1  
D-76344 Eggenstein-Leopoldshafen  
[germaine.goetzelmann@kit.edu](mailto:germaine.goetzelmann@kit.edu)



Philipp Hegel  
Freie Universität Berlin  
Sonderforschungsbereich 980  
Schwendenerstraße 8  
D-14195 Berlin  
[philipp.hegel@fu-berlin.de](mailto:philipp.hegel@fu-berlin.de)



Dr. Michael Krewet  
Freie Universität Berlin  
Sonderforschungsbereich 980  
Schwendenerstraße 8  
D-14195 Berlin  
m.krewet@fu-berlin.de



Sibylle Söring  
Freie Universität Berlin  
Center für Digitale Systeme (CeDiS)  
Garystraße 39  
D-14195 Berlin  
ssoering@cedis.fu-berlin.de



Dr. Danah Tonne  
Karlsruher Institut für Technologie  
Steinbuch Centre for Computing, Hermann-von-Helmholtz-Platz 1  
D-76344 Eggenstein-Leopoldshafen  
danah.tonne@kit.edu